International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

A Quantitative Real-Time Analysis Of Object Tracking Algorithm For Surveillance Applications

J. Arunnehru[1], M. Kalaiselvi Geetha[2]

[1](Research Scholar) Computer Science and Engineering, Annamalai University, Chidambaram, India;
[2](Associate Professor) Computer Science and Engineering, Annamalai University, Chidambaram, India.

Email: arunnehru.aucse@gmail.com, geesiv@gmail.com

Abstract

Three major object tracking algorithms are evaluated in this paper. The performances of these algorithms for different video sequences are analysed. Object tracking is an important task in many surveillance applications, some problem and its difficulty depend on several factors such as pose change, various lighting condition, occlusion, dynamic object, scale and object motion, recent tracking algorithm focus on these three things robustness, adaptively and real-time processing. This paper analyses and identifies the major roles of tracking methods to meet their particular challenges and suggest how to design and choose the tracking algorithm for various conditions. Multiple instance learning tracker (MILT), visual tracking decomposition algorithm (VTD) and track-learning-detection method (TLD) algorithms are tested on various challenging sequence to obtain comparative results. Strength and weakness of these algorithms according to the degree of evaluation criteria are also analysed.

***Keywords*- Multiple instance learning tracker; visual tracking decomposition; Track-learning-detection; Object Tracking; Video Surveillance**

I. INTRODUCTION

In recent years extensive investigation and analyses have been done in the domain of object tracking, tracking of moving objects in video processing, This plays a very important role in many vision based applications like human computer interaction [1], traffic pattern analysis [2] and surveillance based application, The main aim of object tracking is to find location parameters and motion in the image sequence, several studies of the object tracking proposed numerous tracking algorithm for various tasks, It is a challenging problem to implement one tracking algorithm for all task and situations involving different orientation and object poses, This paper, analyses recent tracking methods and evaluate their performance on different datasets.

A typical tracking system consists of three components: object representation, dynamic model, and search mechanism. Object representation is a key component as it directly corresponds to the core challenge of tracking, i.e., how to match object appearance despite all the influencing factors to determine and objective function can be used for searching the target frames. To deal with the problem of appearance variations, recent tracking algorithms focus on adaptive object representation schemes based on generative or discriminative formulations.

A dynamic model, either predefined or learned from certain training data, is often used to predict the possible target states (e.g., motion parameters) in order to reduce the search space and computational load. Since it is difficult to adapt an effective dynamic model for fast motion and as a result of faster processors, most current tracking algorithms use random walk model to predict the likely states.

To distinguish the target region from background using discriminative methods pose object tracking as binary classification, Grabner et al. [3] built a tracking method, which selects features from appearance variations of the object caused by various lighting conditions and out-of-plane rotations. Avidan [4] support vector machine (SVM) is used to trains the classifier and combines it with optical flow for object tracking. Collins et al. [5] propose a tracking method to track the object using selective discriminative low-level color features. whereas Avidan [6] uses a boosting method to classify pixels belonging to foreground and background. Kalal et al. [7] treat sampled data during tracking as unlabeled ones and exploit their underlying structure to select positive and negative samples for update. Babenko et al. [8] use multiple instances learning (MIL) to update classifier to handle obscurely labeled positive and negative data to reduce visual drift.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

This paper focuses on evaluating the most recent tracking algorithms multiple instance learning tracker (MILT), visual tracking decomposition algorithm (VTD) and track-learning-detection method (TLD) algorithms are dealing with different challenges, on the other hand, these methods are often evaluated with a few datasets and it is obvious to choose which algorithm should be used for surveillance applications. Several measures such as rate of success and center location error have been used for performance evaluation in tracking algorithm. Initially this paper, deals with adaptive models is critical to deal with the predictable appearance change of the target and background over time. Next, this approach, evaluate tracking algorithms to identify their key roles in different challenging cases which is a mentioned above, This analysis and evaluation are helpful to built new tracking algorithms and useful for choosing suitable methods for surveillance applications.

II. Challenges In Object Tracking

A major challenge in object tracking is change of shapes of the target object due to the various lighting condition and camera view position, and other difficulties to track the objects are blurred motion and occlusion, some algorithms can deal with the problems of abrupt appearance change, leaving out from scenes and drifting, etc. In some applications with partial lighting condition, static appearance models based on SIFT, [9] HOG [10] and LBP [11] descriptors may be enough. The tracking results may affect due to occlusion in some cases, a dynamic model facilitates reinitialization of a tracker after partial or full occlusions to reduce the search space of states. When a tracking algorithm is designed to account for in-plane motion and scale change with the similarity transform, a static appearance model may be an appropriate option. **Table 1** summarizes challenges and solutions under various conditions in object tracking. To build a robust tracking system, some requirements should be considered. *Robustness:* Robustness means that even under complicated conditions, the tracking algorithms should be able to follow the interested object. The tracking difficulties may be cluttered background, partial and full changing illuminations, occlusions or complex object motion. *Adaptively:* Additional to various changes of the environment that an object is located in, the object itself also undergoes changes. This requires a steady adaptation mechanism of the tracking system to the actual object appearance. *Real-time processing*: A system that needs to deal with live video streams must have high processing speed.

Thus, a fast and optimized implementation as well as the selection of high performance algorithms is required. The processing speed depends on the speed of the observed object.

III. Key Components Of Tracking Algorithm

3.1. Object Representation

In a tracking scenario, an object can be defined as anything that is of interest for further analysis. For instance, vehicles on a road, people walking on a road are a set of objects that may be important to track in a specific domain. Objects can be represented by their shapes and appearances, an object can be represented by raw pixel and color histograms values are called as local descriptors, Color histograms have been used in the particle based method [12] and the mean-shift tracking algorithm [13]. The advantages of histogram-based representations are their computational efficiency and effectiveness to handle shape deformation as well as partial occlusion. The problem discussed in this paper [14, 15] structural information of the target object in histogram-based representation is not designed to handle the scale changes. Local descriptors have also been widely used in object tracking recently due to their robustness to pose and illumination change. Local histograms and color information are utilized for generating confidence maps from which likely target locations can be determined [16].

The meanshift algorithm is a non-parametric method [17]. It provides accurate localization and efficient matching without expensive exhaustive search. The size of the window of search is fixed. It is an iterative process, that is to say, first compute the meanshift value for the current point position, then move the point to its meanshift value as the new position, then compute the meanshift until it fulfill certain condition. For a frame, the distribution of the levels of grey which gives the description of the shape and centre of mass of the object calculated by means of moments [18].

3.1.1 Adaptive Appearance Model

Some of the tracking algorithms are eventually fail to locate the targets, to reduce the visual drifts errors in tracking algorithm, several algorithm have been developed to facilitate adaptive appearance models, It is difficult to determine tracking results in significant drifts, whether the new data are noisy or not. Therefore, drifting errors are likely to collect slowly and tracking algorithms ultimately fail to find the targets.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

Matthews et al [19] propose a tracking method with the Lucas-Kanade algorithm by using fixed reference template is extracted from the first frame to update the template in most recent frames to acquire the results. In contrast to supervised discriminative object tracking. Kalal et al [7] also pose the tracking problem as a semi-supervised learning task and exploit the underlying structure of the unlabeled data to select positive and negative samples for update. Babenko et al [8] multiple instance learning (MIL) frameworks obscurely labeled positive and negative data's are obtained to reduce visual drifts in pose tracking problem. Grabner et al [3] semi-supervised task is used for update problem, to reduce drifts both labeled and unlabeled data is used to update the classifier.

3.1.2. Motion Model

Motion models are abstract representations of human motion which execute atomic tasks. The idea is that every motion belongs to a certain motion model, e.g. walk, run, wave with hands, and throw. Motion models are able to produce a given motion in different styles. The most commonly adopted models are translational motion (2 parameters), similarity transform (4 parameters), and affine transform (6 parameters). The tracking methods [20, 21] account for affine transformation of objects between two consecutive frames. If an algorithm is designed to handle translational movements, the tracking results would not be accurate when the objects undergo rotational motion or scale change even if an adaptive appearance model is utilized. The classic Kanade-Lucas-Tomasi algorithm [22] is designed to estimate object locations although it can be extended to account for affine motion [23]. The mean-shift based tracking algorithm [29] is not equipped to deal with scale change or in-plane rotation since the objective function is not differentiable with respect to these motion parameters.

3.2. Dynamic Model

Dynamic model is used to reduce computational complexity in object tracking as describe as state transition i.e., $p(\mathbf{X}_t \mid \mathbf{X}_{t-1})$, between two consecutive frame where $\mathbf{X}_t$ is the state vector at time t. The early tracking methods such as Kalman filter-based trackers are used in Constant velocity and constant acceleration models. The state transition is modeled by a Gaussian distribution, $p(\mathbf{X}_t \mid \mathbf{X}_{t-1}) = N(\Phi_{t-1} \mathbf{X}_{t-1}, \mathbf{X}_{t-1})$, where $\mathbf{\Phi}_{t-1}$ and ξ_{t-1} are the transfer matrix and noise at time $t\text{-}1$, respectively. The recent tracking algorithms adopt random walk models [21, 25] with particle filters to assume the constant velocity or acceleration.

3.3. Search Mechanism

In this survey, object tracking depends mainly on search strategy state to formulate as an optimization problem, search state utilize either deterministic or stochastic methods, If the objective function is differentiable with respect to the motion parameters, then gradient descent methods can be used [23, 22, 24]. Exhaustive search methods are able to achieve good tracking performance at the expense of very high computational load, and thus seldom used in tracking tasks. Sampling-based search methods can achieve good tracking performance when the state variables do not change drastically. Consequently, stochastic search algorithms such as particle filters are trade-offs between these two extremes, with the ability to escape from local minimum without high computational load. Otherwise, either sampling [8, 3] or stochastic methods [21, 25] can be used. Deterministic methods based on gradient descent are usually computationally efficient, but suffer from the local minimum problems.

IJETAE
EXPLORING RESEARCH AND INNOVATIONS

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

Table 1
Challenges and Solutions under various conditions

Challenges	Solution
Occluded partially	The tracker may be reinitialized using sampling of the state space when the target reappears; Apply parts-based model which are not sensitive to occlude partially
Different angles	Adjust various shapes to such appearance change; Apply state model for the different angles relatively
Complete occlusion	The tracker can be reinitialized by search the state space exhaustively when target reappears
Dynamic background/ Dynamic object	Seek a large region of the state space; Complicated dynamic model
Various lighting conditions	Adjust the appearance model to various lighting changes relatively; Apply descriptors which are not sensitive to various lighting condition
Various shapes due to camera position	Adjust the various shapes to camera position relatively; Apply various shapes which is not sensitive to camera position

IV. OBJECT TRACKING ALGORITHMS

Object tracking is an major task within the field of computer vision. Tracking can be defined as the predicament of estimating the path of an object in the image plane as it moves around a scene. In other words, a tracker assigns dependable labels to the tracked objects in various frames of a video. Object tracking is a well-known problem in video surveillance community. Many researchers have addressed the problem in real-world scenarios rather than a lab environment. It is a very challenging task to track an object in appearance changes cover geometric and photometric variations of an object such as occlusion, pose, or various lighting condition changes. To deal with all these changes simultaneously, tracking methods need more complex observation and motion model as well as an efficient tracking model. In this section, three recent algorithms such as MILT, VTD and TLD are discussed.

4.1. Multiple Instance Learning Tracking (MILT)

Multiple instance learning racking (MILT) [8] to handle obscurely labeled positive and negative data's obtained to reduce visual drift caused by classifier update. Haar-like wavelets are used to illustrate objects for boosting based tracking methods and filter responses are used to represent objects.

4.2. Visual Tracking Decomposition (VTD)

Visual tracking decomposition algorithm [26] use conventional particle filter framework with multiple dynamic and observation models for appearance variation and stochastic search algorithms such as particle filters are trade-offs between these two extremes, with the ability to escape from local minimum without high computational load and it is widely used in recent object tracking algorithm.

4.3. Track Learning Detection (TLD)

Tracking learning detection method [7] sampled data are treated as unlabeled ones during tracking and exploit their underlying structure to select positive and negative samples for update and pose the tracking problem as a semi-supervised learning task and exploit the underlying structure of the unlabeled data to select positive and negative samples for update.

V. EXPERIMENTAL EVALUATION OF TRACKING ALGORITHMS

In this section, experiments on three publicly available video sequences, this video has been adopted by many researchers to test their algorithms as well as two of our own, because of its capacity in simulating various tracking conditions, including various lighting conditions, pose variations, occlusions, and distraction.

Analyses the tracking methods for surveillance applications based on above discussion and suggest how to choose and design algorithm effectively, to evaluate 3 recent tracking algorithms on 5 different testing sequences, The test algorithms include: multiple instance learning tracker (MILT), [8] visual tracking decomposition algorithm (VTD), [26] and track learning detection method (TLD), [7].

TLD algorithm uses similarity transform as the motion model. Raw pixels and color histogram are used to describe the objects. MILT algorithm uses translational motion model.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

Raw pixel and color histogram are used to describe the objects. VTD algorithm uses similarity transform as the motion model. Raw pixels and color histogram are used to describe the objects.

A more detailed description of these algorithms is given in **Table 2** and some of the tracking results on various challenges are shown in **Fig 1**.

Table 2.
Tracking Algorithm

Tracking Algorithm	Motion Model	Object Description	Dynamic Model	Searching methods	Properties
Track Learning Detection Method (TLD)	Similarity Transform	Raw pixels and color histogram based on Haar-like descriptor	-	Sampling	Discriminative
Multiple Instance Learning Tracker (MILT)	Translational Motion	Raw pixels and color histogram based on Haar-like descriptor	-	Sampling	Discriminative
Visual Tracking Decomposition Algorithm (VTD)	Similarity Transform	Raw pixels and color histogram on hue, saturation, intensity, and edge template	Gaussian	Particle filter	Generative

Experiments were carried out using 5 different datasets PETS2009, car, indoor, face occlusion, outdoor, which pose various challenging factors to analyse, these video are collected at different resolution. A dataset used for experimental analysis is shown in **Table 3**.

Table 3.
Dataset used in the current analysis

Datasets	Challenges	Resolution	No. of Frames
PETS 2009	pose change, similar objects distraction, full occlusion,	768 x 576	824
car	occlusion partial, blurred image	320 x 240	320
indoor	pose change, occlusion partial, Various lighting conditions	320 x 240	310
face occlusion	occlusion partial, in-plane pose change	320 x 240	920
outdoor	low-contrast, pose change, occlusion,	320 x 240	340

The quantitative evaluation are used to find the rate of success and location error with the reference of object centre, and employ the method PASCAL VOC [27] used to compute the rate of success and score is defined as given below in **Eq. 1.**

$$score = \frac{area(ROI_T \cap ROI_G)}{area(ROI_T \cup ROI_G)} \qquad (1)$$

Where: ROI_T is the tracked bounding box
ROI_G is the ground truth bounding box

The rate of success is computed in all frames, where this score > 0.5 in each frame is considered as a success is shown in **Table 4**.

Table 4.
Rate of Success (%)

Datasets	MILT	VTD	TLD
PETS2009	40	25	80
car	80	87	92
indoor	53	80	71
face occlusion	85	62	85
outdoor	42	53	35

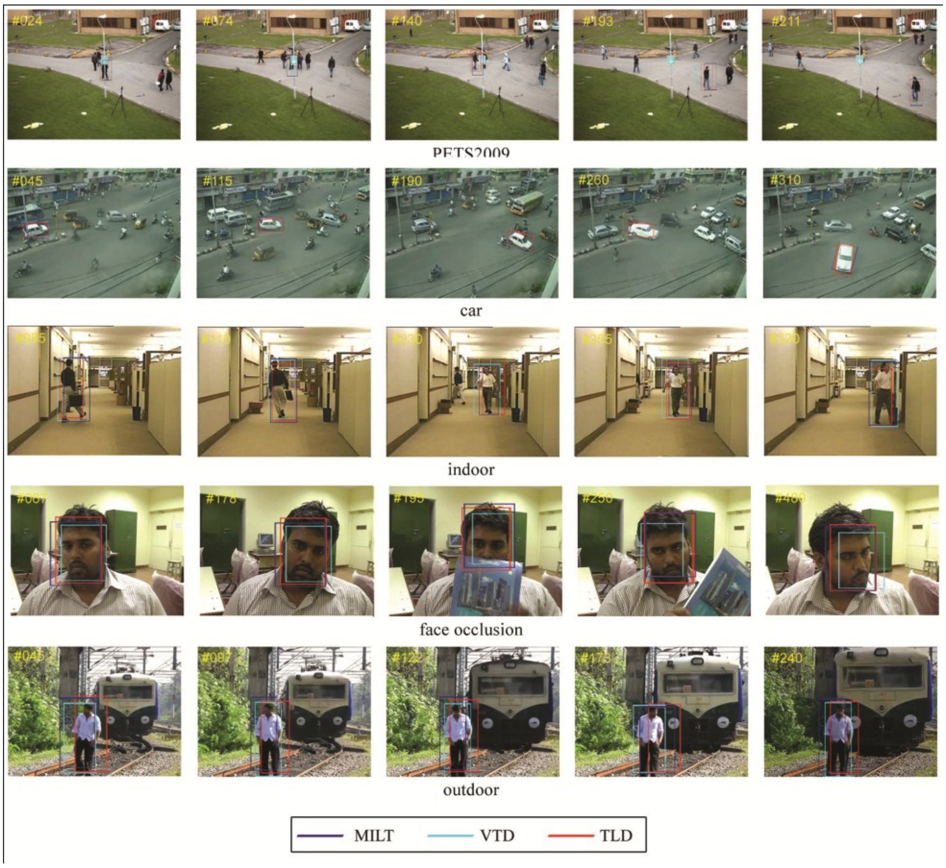

Fig. 1. Tracking analysis results on various challenges

The centre location error rate is computed in all frames, where this score < 0.5 in each frame is considered as an error rate is shown in **Table 5**.

Table 5
Centre location error rate (%)

Datasets	MILT	VTD	TLD
PETS2009	60	75	20
car	20	13	8
indoor	47	20	29
face occlusion	15	38	15
outdoor	58	47	65

The MILT method uses to update the appearance model in multiple instances learning to decrease the visual drifts. It performs well in car and face occlusion.

The MILT method is not deliberate to handle pose change, low contrast and scale changes, it does not make fine in the target sequences which undergo large scale changes and pose change (e.g., outdoor, PETS2009), The VTD method utilize observation model and multiple dynamics for various shapes due to change of camera position, it performs fine indoor and car sequences, where some distracters occurred in the scene due to various lighting condition in outdoor and PETS2009. The TLD method works superior than the other methods in the PETS2009, car, face occlusion, when change in shapes (e.g., when the subject place book in front of his face, or turns his face), it not perform well in outdoor and indoor sequences due to distracters objects are similar to target objects in clustered environment. It is immense significance to design more discriminative appearance models with update model are used in clustered environment to split target objects without any tracking errors.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

V. CONCLUSION

This paper analyses 3 recent tracking algorithms with different challenging sequence. This paper also evaluates the roles of object tracking with complete analysis on the performance in surveillance applications. The analysis shows the robustness and weakness of these tracking algorithms, which helps to choose and design an effective tracking algorithm that is suitable under various conditions.

Acknowledgements

This work is being supported by University Grants Commission of India [F. No. 41-636/2012 (SR)].

REFERENCES

[1] M. de La Gorce, N. Paragios, and D. Fleet. 2008. Model-based hand tracking with texture, shading and self-occlusions, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.

[2] Z. Sun, G. Bebis, and R. Miller. 2006. On-road vehicle detection: A review, IEEE Transactions on Pattern Analysis and Machine Intelligence 28(5), pp. 694–711.

[3] H. Grabner, C. Leistner, and H. Bischof. 2008. Semi-supervised on-line boosting for robust tracking, Proceedings of European Conference on Computer Vision, pp. 234–247.

[4] S. Avidan. 2004. Support vector tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence 26(8), pp. 1064–1072.

[5] R. T. Collins, Y. Liu, and M. Leordeanu. 2005. Online selection of discriminative tracking features, IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), pp. 1631–1643.

[6] S. Avidan. 2007. Ensemble tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence 29(2), pp. 261–271.

[7] Z. Kalal, J. Matas, and K. Mikolajczyk. 2010. P-n learning: Bootstrapping binary classifiers by structural constraints, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 49–56.

[8] B. Babenko, M.-H. Yang, and S. Belongie. 2009. Visual tracking with online multiple instance learning, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 983–990.

[9] D. Lowe. 2004. Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60(2), pp. 91–110.

[10] N. Dalal and B. Triggs. 2005. Histograms of oriented gradients for human detection, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.

[11] T. Ojala, M. Pietikainen, and T. Maenpaa. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), pp. 971–987.

[12] P. Pierez, C. Hue, J. Vermaak, and M. Gangnet. 2002. Color-based probabilistic tracking, Proceedings of European Conference on Computer Vision, pp. 661–675.

[13] D. Comaniciu, V. Ramesh, and P. Meer. 2003. Kernel-based object tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence 25(5), pp. 564–575.

[14] R. T. Collins. 2003. Mean-shift blob tracking through scale space, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 234–240.

[15] S. T. Birchfield and S. Rangarajan. 2005. Spatiograms versus histograms for region-based tracking, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2, pp. 1158–1163, 20–25.

[16] S. Avidan. 2007. Ensemble tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence 29(2), pp. 261–271.

[17] S. Franois, B. and R.J. Alexandre. 2004. Camshift Tracker Design Experiments. IMSC, no. 11, pp. 1–11.

[18] G. Bradski, and T. Ogiuchi, and M. Higashikubo. 2010. Visual Tracking Algorithm using Pixel-Pair Feature, International Conference on Pattern Recognition, no. 4, pp. 1808–1811.

[19] L. Matthews, T. Ishikawa, and S. Baker. 2004. The template update problem, IEEE Transactions on Pattern Analysis and Machine Intelligence 26(6), pp. 810–815.

[20] X. Li, W. Hu, Z. Zhang, X. Zhang, and G. Luo. 2007. Robust visual tracking based on incremental tensor subspace learning, Proceedings of the IEEE International Conference on Computer Vision.

[21] X. Mei and H. Ling. 2009. Robust visual tracking using l1 minimization, Proceedings of the IEEE International Conference on Computer Vision, pp. 1436–1443.

[22] B. Lucas and T. Kanade. 1981. An iterative image registration technique with an application to stereo vision, Proceedings of International Joint Conference on Artificial Intelligence, pp. 674–679.

[23] L. Matthews, T. Ishikawa, and S. Baker. 2004. The template update problem, IEEE Transactions on Pattern Analysis and Machine Intelligence 26(6), pp. 810–815.

[24] D. Comaniciu, V. Ramesh, and P. Meer. 2003. Kernel-based object tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence 25(5), pp. 564–575.

[25] D. Ross, J. Lim, R.-S. Lin, and M.-H. Yang. 2008. Incremental learning for robust visual tracking, International Journal of Computer Vision 77(1-3), pp. 125–141.

[26] J. Kwon and K. Lee. 2010. Visual tracking decomposition, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1269–1276.

[27] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman. 2010. The pascal visual object classes (voc) challenge, International Journal of Computer Vision 88(2), pp. 303–338.